

Big Data: not just a lot more data

Darrin Baines MSc, PhD

For over two decades, the NHS has had easy access to accurate and plentiful prescribing data. Since the introduction of the Prescribing Analysis and Cost (PACT) database in the late-1980s, millions of pieces of information have been routinely produced and analysed on doctor drug choice. The announcement of the impending arrival of a 'Big Data' revolution may therefore seem like old news to those involved in prescribing analysis.

However, Big Data promises to be more than just an increase in the amount of information available. This new paradigm intends to change the ways in which prescribing data are collated, processed and analysed, with greater benefits for patients.

What is Big Data?

There are a number of different definitions of Big Data, but the term has its origins in a research paper written in 2001 by Doug Laney.¹ Outlining the future challenges for informatics management, Laney saw the threats ahead as:

- an unprecedented growth in data volumes
- a rapid increase in the velocity or speed of data creation
- an extension in the variety of data types available.

Therefore Big Data is not just lots and lots of information – it is also concerned with the advancing speed in which new information, from an increasing number of diverse sources, is confronting evolving organisations. In other words, Big Data is about the dangers of being overloaded with information, arriving at a quickening pace from previously unknown directions. In the future, not only will we see greater computing capacity for producing NHS drug data, but this information will be routinely combined with an assortment of information from varying sources.

For instance, Hospital Episode Statistics (HES) will probably be available alongside Prescribing Cost Analysis (PCA) data streams. There will therefore be more information related to prescribing involving what is called a 'mash-up' of sources.

Although published over a decade ago, the predictions in Laney's paper are still proving correct: most data-collecting entities are experiencing a sustained growth in volumes, velocity and variety, which constantly challenges their approach to information management. For many in the informatics industry, this situation suggests that organisations need Big Data solutions to address their Big Data problems.

In the main, opinion leaders in the field are looking for technical, computing-based answers to what they view as a situation stimulated by rapid developments in computer software and hardware. As data streams are becoming so large, fast flowing and diverse in their sources, better computing programming, innovative analytical algorithms and more sophisticated architecture for storage are seen as the common solutions to the expanding problem.

For the NHS, larger and faster data sources are going to challenge the capacity of the existing system to analyse prescribing data, which is led by information analysts and clinical advisers.

The Big Data revolution will therefore require local bodies like clinical commissioning groups (CCGs) to ask whether they have the right structures in place to analyse the data available to them. Big Data will not only revolutionise the knowledge available at our fingertips but may also create new approaches to how analysts and clinicians work.

Understanding the potential

Although there is a constant buzz about the potential of Big Data to change the world, there is always the possibility that the noise being made is just hype without substance. However, in his book *Disruptive Possibilities: How Big Data Changes Everything*,² Jeff Needham states, 'Big Data is not just this season's trendy hemline, it is a must-have piece that will last for generations'.

He suggests that this technology is a type of supercomputing for commercial enterprises and governments that will make it possible to monitor pandemics as they happen, to anticipate where the next bank robbery will occur, to predict voter behaviour and to forecast the volatility of political uprisings while they happen. The revolutionary potential of this new technology will therefore create world-changing possibilities that will transform business and government forever.

Given its disruptive potential, academics, decision-makers, prescribers and pharma will need to explore the possible impact of this new paradigm on information management in the NHS, including the processing and production of data on prescribing.

Improving analysis

For people interested in the area but not versed in the terminology of data management and supercomputing, *Big Data for Dummies*³ is a useful read. The author reports that, 'Big Data is not a single technology but a combination of old and new technologies that helps companies gain actionable insight'. Therefore, Big Data is the 'capability to manage a huge volume of disparate data, at the right speed and within the right time frame, to allow real-time analysis and reaction'. This definition is in keeping with the notions of volume, velocity and variety discussed above.

However, the definition suggests that Big Data has the capability of doing much more than producing numbers for managers to digest. The switch to Big Data offers innovative opportunities to improve the ways in which we analyse health service information, which should lead to marked improvements in patient care and health outcomes.

With the appointment of one of the co-founders of Dr Foster, Tim Kelsey, to the role of National Director for Patients and Information, the NHS has signaled its seriousness to adopt health service informatics. Last year he also became the government's Executive Director of Transparency and Open Data. This appointment suggests that, in the future, large amounts of NHS data will be available in the public domain for both public and private entities to access, process and analyse. Prescribing data will thus soon be part of a growing Big Data agenda for the NHS. However, this begs the question of whether enough of us yet understand the Big Data agenda well enough to map out the path ahead.

References

1. Laney, D. *3D data management: controlling data volume, velocity and variety*. META Group Inc, 2001.
2. Needham, J. *Disruptive possibilities: how big data changes everything*. O'Reilly, 2013.
3. Hurwitz, J. *Big data for dummies*. John Wiley & Sons, 2013.

Declaration of interests

None to declare.

Darrin Baines is associate professor, Faculty of Medicine & Health Sciences, University of Nottingham